



Heriot-Watt University
Research Gateway

New model diagnostics for spatio-temporal systems in epidemiology and ecology

Citation for published version:

Lau, MSY, Marion, G, Streftaris, G & Gibson, GJ 2014, 'New model diagnostics for spatio-temporal systems in epidemiology and ecology', *Journal of the Royal Society. Interface*, vol. 11, no. 93, 20131093.
<https://doi.org/10.1098/rsif.2013.1093>

Digital Object Identifier (DOI):

[10.1098/rsif.2013.1093](https://doi.org/10.1098/rsif.2013.1093)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Journal of the Royal Society. Interface

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

New model diagnostics for spatio-temporal systems in epidemiology and ecology

Max S.Y. Lau^{1,3}, Glenn Marion³, George Streftaris^{1,2}, and Gavin J. Gibson^{1,2}

¹Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, United Kingdom, EH14 4AS

²Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, United Kingdom, EH14 4AS

³Biomathematics and Statistics Scotland, Edinburgh, United Kingdom, EH9 3JZ

Abstract

A cardinal challenge in epidemiological and ecological modelling is to develop effective and easily deployed tools for model assessment. The availability of such methods would greatly improve understanding, prediction and management of disease and ecosystems. Conventional Bayesian model assessment tools such as Bayes factors and the Deviance Information Criterion (DIC) are natural candidates but suffer from important limitations due to their sensitivity and complexity. Posterior predictive checks, which utilize summary statistics of the observed process simulated from competing models, can provide a measure of model fit but appropriate statistics can be difficult to identify. Here we develop a novel approach for diagnosing mis-specifications of a general spatio-temporal transmission model by embedding classical ideas within a Bayesian analysis. Specifically, by proposing suitably designed non-centered parameterization schemes, we construct latent residuals whose sampling properties are known given the model specification and which can be used to measure overall fit and to elicit evidence of the nature of mis-specifications of spatial and temporal processes included in the model. This model assessment approach can readily be implemented as an addendum to standard estimation algorithms for sampling from the posterior distributions such as Markov chain Monte Carlo. The proposed methodology is first tested using simulated data and subsequently applied to data describing the spread of *Heracleum mantegazzianum* (giant hogweed) across Great Britain over a thirty-year period. The proposed methods are compared with alternative techniques including posterior predictive checking and the DIC. Results show that the proposed diagnostic tools are effective in assessing competing stochastic spatio-temporal transmission models and may offer improvements in power to detect model mis-specifications. Moreover, the latent-residual framework introduced here extends readily to a broad range of ecological and epidemiological models.

Keywords: Spatio-temporal model assessment | Latent residuals | Non-centered parameterization | Bayesian Inference

1 Introduction

Stochastic spatio-temporal models are playing an increasingly important role in epidemiological and ecological studies relating to transmission of diseases [1, 2], invasion of alien species [3] and population movements in response to climate changes [4]. It is well known that the predicted dynamics of such systems can be extremely sensitive to the choice of model, with consequent implications for the design of control strategies [1, 5, 6, 7], but as yet there is a lack of effective model assessment tools described in the literature. For example, studies of foot and mouth disease have cited the importance of selecting between a long-tailed *spatial kernel* (see Section 2) versus a localized spatial kernel, but this model-choice problem is far from being resolved [8, 5]. Further model-choice problems arise in relation to the parametric form of the distributions of incubation and infectious periods in models of measles [9, 10, 11], and in relation to diseases such as smallpox [12] and AIDS [13].

Bayesian model assessment techniques appear appealing [14, 15, 16] particularly since many of the above studies use Bayesian techniques for model fitting. However it is well known that this approach is extremely sensitive to prior

assumptions regarding the distribution of parameters within the competing models. Moreover, the computational challenges presented by Bayesian model choice algorithms can be prohibitive. An alternative approach, based on the *Deviance Information Criterion (DIC)* [17], is known to be problematic when applied to processes that are only partially observed [18], as in the case of most real-world systems. Furthermore, neither of these approaches leads to a diagnostic tool for assessing the fit of a model in absolute terms; rather they provide an assessment of the *relative* fit of the competing models. Posterior predictive checks [19], whereby summary statistics of the observed process are compared to their predictive distributions, do provide a measure but appropriate statistics can be difficult to identify [20].

We address this gap in available methodology by pursuing the following objectives:

- to innovate a statistically sound framework for assessing stochastic spatio-temporal models, which can be readily implemented as an addendum to a Bayesian analysis and which avoids the sensitivity and complexity of Bayesian model assessment;
- to illustrate how the approach can be targeted to assess particular aspects of a spatio-temporal stochastic model, here principally the choice of spatial kernel;
- to demonstrate the effectiveness of the approach using simulated data and to apply it to an ecological dataset describing the spread of an alien species throughout Great Britain.

The approach adopted (see Section 2) involves representing stochastic spatio-temporal models using appropriately designed *non-centered parameterization* schemes [21], from which *latent residual processes* can be defined. The assessment of fit of a model to a given data set is then achieved in the Bayesian framework by imputing these residuals, and testing them for compliance with their (known) sampling model using classical tests. The approach has its roots in the framework proposed in [19], and extended in [22] and [20]. The key innovation in this paper is to design the residual processes so that the resulting tests are sensitive to mis-specification of specific aspects of the model under consideration. In particular, we formulate tests for detecting mis-specification of the spatial transmission kernel as this aspect typically has major implications for control strategies, for example based on culling or removal of susceptibles in the neighbourhood of infected individuals.

2 Model and Methods

2.1 Spatio-temporal stochastic model

We consider a broad class of spatio-temporal stochastic models exemplified by the SEIR epidemic model with susceptible (S), exposed (E), infectious (I) and removed (R) compartments. Suppose that we have a spatially distributed population indexed 1, 2, ..., denote $\xi_S(t)$, $\xi_E(t)$, $\xi_I(t)$ and $\xi_R(t)$ as the set of indices for individuals who are in class S, E, I and class R respectively at time t , and let $S(t)$, $E(t)$, $I(t)$ and $R(t)$ be the respective numbers of individuals in these classes at time t . Then individual $j \in \xi_S(t)$ becomes exposed during $[t, t + dt)$ with probability

$$p(j, t) = \{\alpha + \beta \sum_{i \in \xi_I(t)} K(d_{ij}, \kappa)\} dt + o(dt), \quad (1)$$

where α represents a primary infection rate and β is a contact parameter. The term $K(d_{ij}, \kappa)$ characterises the dependence of the infectious challenge from infective i to j as a function of distance d_{ij} and is known as the *spatial kernel function*. Following exposure, the random times spent by individuals in classes E and I are modelled using a suitable distribution such as a Gamma or a Weibull distribution [12, 23, 24]. Specifically, we use a *Gamma*(μ, σ^2) parameterized by the mean, μ , and variance, σ^2 , for the random time x spent in class E with density function

$$f_E(x; \mu, \sigma^2) = \frac{1}{(\sigma^2/\mu)^{\mu^2/\sigma^2} \Gamma(\mu^2/\sigma^2)} x^{\mu^2/\sigma^2 - 1} e^{-x\mu/\sigma^2}. \quad (2)$$

For the random time x spent in class I we use a *Weibull*(γ, η) parameterized by the shape and scale with density function

$$f_I(x; \gamma, \eta) = (\eta/\gamma)(x/\gamma)^{\eta-1} e^{-(x/\gamma)^\eta}. \quad (3)$$

All sojourn times are assumed independent of each other given the model parameters.

2.2 Latent Residuals

Let \mathbf{Z} denote the complete set of data (that is the time, nature and affected individual for of all transitions between states) describing an epidemic generated randomly from the above model parametrized by $\boldsymbol{\theta}$. Then, as long as the sampling properties of the stochastic model described by Equation (1) are preserved, we can consider \mathbf{Z} to be generated in non-unique ways. In particular, we consider \mathbf{Z} as a deterministic function $h_{\boldsymbol{\theta}}(\cdot)$ of a random vector $\tilde{\mathbf{r}}$ where the components of the latter are generated as a random sample from a Uniform(0, 1) distribution. That is

$$\mathbf{Z} = h_{\boldsymbol{\theta}}(\tilde{\mathbf{r}}). \quad (4)$$

This representation is essentially a *functional model* in the sense of [25], and is an illustration of the concept of generalised residuals proposed in [26]. Note that the selection of a residual process $\tilde{\mathbf{r}}$ and a function $h_{\boldsymbol{\theta}}(\cdot)$ which together specify the model given by Equation (1) can be effected in a multiplicity of ways. In Section 2.2.2 we consider how this selection may be done in order to facilitate the design of statistical tests based on $\tilde{\mathbf{r}}$ that can be sensitive to mis-specification of particular aspects of the model.

The particular construction we exploit involves a process $\tilde{\mathbf{r}}$ that can be partitioned into four independent random samples from Uniform(0, 1) and expressed as $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4)$ where each \tilde{r}_j , $j = 1, 2, 3, 4$, is a vector which determines events relating to a different aspect of the process. The process $\tilde{r}_1 = (r_{11}, r_{12}, \dots)$ defines a set of population-level thresholds from which the time of each subsequent exposure can be determined by considering the integrated infectious challenge. For the k^{th} exposure (ordered temporally), r_{3k} and r_{4k} specify the quantile of the random sojourn times in class E and I respectively. The residuals \tilde{r}_2 determine the particular infectious contacts that generate each exposure. We now describe concisely how the epidemic process can be constructed through the residuals \tilde{r}_1 , \tilde{r}_2 and \tilde{r}_3 . A full description of the residuals can be found in the *Electronic Supplementary Material (ESM)*.

2.2.1 Exposure Time Residuals (ETR)

We refer to the residuals \tilde{r}_1 as *Exposure Time Residuals (ETR)*. Starting from the $(k-1)^{th}$ exposure event, we define the accumulated infectious challenge in the population by time t as

$$A(t) = \int_{t_{k-1}}^t \sum_{j \in \xi_S(y)} \{ \alpha + \beta \sum_{i \in \xi_I(y)} K(d_{ij}, \kappa) \} dy$$

where t_{k-1} is the time of the $(k-1)^{th}$ exposure event. The time of k^{th} exposure is then determined from

$$t_k = \inf \left\{ t \mid 1 - e^{-A(t)} > r_{1k} \right\}. \quad (5)$$

2.2.2 Infection-Link Residuals (ILR)

We refer to the residuals \tilde{r}_2 as *Infection-Link Residuals (ILR)*. Given that the k^{th} exposure event occurs during $(t_k, t_k + dt)$, and given the other transitions that have occurred prior to t_k , the probability that the respective contact is between susceptible $j \in \xi_S(t_k)$ and infective $i \in \xi_I(t_k)$, is given by

$$p_{ij} \propto K(d_{ij}, \kappa). \quad (6)$$

Note that the primary infection process can be accommodated by adding a notional and permanently infectious individual which presents a challenge α to every susceptible. To generate the particular infection-link from the residual, r_{2k} , we arrange the p_{ij} in the ascending order $p_{(1)}, \dots, p_{(m)}$ where $m = S(t_k)(I(t_k) + 1)$ is the total number of ‘active’ links. The link responsible for the k^{th} exposure is then determined from

$$s' = \inf \left\{ s \mid \sum_{a=1}^s p_{(a)} > r_{2k} \right\}, \quad (7)$$

so that individual j becomes exposed due to contact with individual i , and $p_{ij} = p_{(s')}$. The inclusion of the ordering operation in this process is motivated by our aim of designing tests that may be sensitive to mis-specification of the spatial kernel function in the model Equation (1). Suppose that the modelled kernel function deviates from reality in a systematic way - for example by declining too rapidly, or too slowly with distance. Then a heuristic argument (see

ESM Section 1.4) suggests that we may see a correspondingly systematic deviation from the $U(0, 1)$ distribution when the residuals \tilde{r}_2 are imputed (as described in Section *reImputation of Infection-Link Residuals (ILR)*) from observations, with a concentration, or scarcity of residuals at the extremes, so that model mis-specification may be readily detected using standard tests of the fit of the imputed \tilde{r}_2 to the uniform distribution.

2.2.3 Latent Time Residuals (LTR)

We refer to the residuals \tilde{r}_3 as *Latent Time Residuals (LTR)*. For the k^{th} exposure, define the accumulated pressure of becoming infectious by time t as

$$Q(t) = \int_{t_k}^t \frac{f(y)}{1 - F(y)} dy, \quad (8)$$

where $f(y)$ and $F(y)$ are to the density and cumulative distribution function of the latent period respectively. The time of becoming infectious is then determined from

$$t'_k = \inf \left\{ t \mid 1 - e^{-Q(t)} > r_{3k} \right\}. \quad (9)$$

We remark that the time of recovery can be determined similarly by utilizing r_{4k} and an appropriate sojourn time distribution in class I .

2.3 Bayesian inference and model assessment

It is now standard practice to conduct Bayesian analyses of partially observed epidemics using the process of *data augmentation* supported by computational techniques such as Markov chain Monte Carlo methods [5, 27, 28, 20]. Given partial data y , these approaches involve simulating from the joint posterior distribution $\pi(\boldsymbol{\theta}, \mathbf{z}|y)$ where \mathbf{z} represents the complete epidemic data as above. As applied to fit models in this paper, is described more fully in the *ESM*. As the complete epidemic \mathbf{z} is reconstructed, or ‘imputed’, it naturally lends itself to the residual-based testing methods now described.

Given a random draw $(\boldsymbol{\theta}', \mathbf{z}')$ from $\pi(\boldsymbol{\theta}, \mathbf{z}|y)$ it is generally straightforward to invert Equation (4) to impute the corresponding residual $\tilde{\mathbf{r}}'$ by sampling it uniformly from the set $h_{\boldsymbol{\theta}'}^{-1}(\mathbf{z}')$ (see Section 2.3.1), the set of residuals mapped to \mathbf{z}' by $h_{\boldsymbol{\theta}'}$. Therefore a sample from the posterior distribution $\pi(\tilde{\mathbf{r}}|y)$ can easily be generated with a minor insert to an existing algorithm. On applying a classical test for consistency with the uniform distribution to the $\tilde{\mathbf{r}}'$ (specifically we use an *Anderson-Darling* hypothesis test [29], see also *ESM*) a posterior distribution of *p-values*, $\pi(P(\tilde{\mathbf{r}})|y)$ is generated, from which evidence against the modelling assumptions can be discerned. In Bayesian parlance we note that the pair $(\boldsymbol{\theta}', \tilde{\mathbf{r}}')$ represents a *non-centered parameterization*.

2.3.1 Imputation of Infection-Link Residuals (ILR)

The imputation of \tilde{r}_1 and \tilde{r}_3 given $(\boldsymbol{\theta}', \mathbf{z}')$ is straightforward by inverting the procedure specified by Equation (5) and Equation (9) respectively. Imputation of r_{2k} is achieved by inverting Equation (7) but, since the infection link is discrete and the space of residuals continuous, the imputation process warrants description here. The particular infection link for the k^{th} exposure event is randomly chosen from the links between the corresponding exposed individual k and $i \in \xi_I(t)$ according to probabilities p_{ik} defined in Equation (6). The ranking of this particular infection link, s' , is then determined among all links between $j \in \xi_S(t)$ and infective $i \in \xi_I(t)$. Finally, the residual r'_{2k} is imputed as a random draw

$$r'_{2k} \sim U \left(\sum_{i=1}^{i < s'} p_{(i)}, \sum_{i=1}^{i < s'} p_{(i)} + p_{(s')} \right). \quad (10)$$

Bayesian residuals have been utilized in other contexts [30]. In [20] it was shown that Bayesian latent residuals based on Sellke thresholds [31] could be used to assess the fit of spatio-temporal models. However the specific approach is problematic when the epidemic is small as thresholds must be imputed even for uninfected individuals. By contrast the construction proposed here requires residuals to be imputed only for each infection event and avoids this shortcoming. Moreover, since the components of the residual process $\tilde{\mathbf{r}}$ each relate to a different aspect of the stochastic model, then it is plausible that testing for mis-specification of a given aspect may be best achieved by considering only the relevant component of $\tilde{\mathbf{r}}$. In particular, mis-specification of a spatial kernel or the latent period distribution may be assessed by examining the posterior samples of \tilde{r}_2 (ILR), and \tilde{r}_3 (LTR) respectively. We stress again the importance for the detection of mis-specified spatial kernels of the ordering operation (see Section 2.2.2)

in the construction of the ILR, which is included with the expectation that it leads to systematic, detectable, and interpretable deviations from $U(0, 1)$ in the imputed residuals. This issue is discussed further in *ESM (Section 1.4)* and Section 3.1 below. As described in Section 3 and Section 4, this hoped-for sensitivity is indeed achieved.

2.4 Interpretation of latent-residual tests

Posterior distributions of p-values arising from a classical test applied to a latent process have been exploited in [20, 22, 32]. For completeness we include some comments on the statistical interpretation of such distributions in the Bayesian context. To the Bayesian observer of data y , $\pi(P(\tilde{\mathbf{r}})|y)$ represents their posterior belief regarding the p-value that a classical observer of $\tilde{\mathbf{r}}$ would compute. Should this distribution be concentrated on small values, the Bayesian would infer that the classical observer may reject the hypothesis that the $\tilde{\mathbf{r}}$ were generated as a random sample from a $U(0, 1)$ distribution. The latter hypothesis is a key assumption for the functional-model representation given in Equation (4) so that the classical observer would likewise question the validity of this model. Therefore, the Bayesian observer can extract from $\pi(P(\tilde{\mathbf{r}})|y)$ summaries such as $\pi(P(\tilde{\mathbf{r}}) < 0.05|y)$ (as used here) and interpret them as measures of evidence against their model assumptions.

3 Simulated Example

To test the methodology we apply it to analyse spatio-temporal epidemics simulated in a population of size $N = 1000$, whose locations are sampled independently from a uniform distribution over a square region, between times $t = 0$ and $t = t_{max} = 50$. We assume that the entire population is susceptible at time 0, that the epidemic evolves according to Equation (1) with $\alpha = 0.001$, $\beta = 3$, $K(d_{ij}, \kappa_1) = \exp(-0.03d_{ij})$, and that the sojourn times in classes E and I follow *Gamma*(5, 2.5) and *Weibull*(2, 2) distributions respectively. The observations y constitute only the precise times and locations of transitions from E to I and from I to R that occur during the observation period. Figure 1 illustrates the spatio-temporal progression of a typical realisation of y .

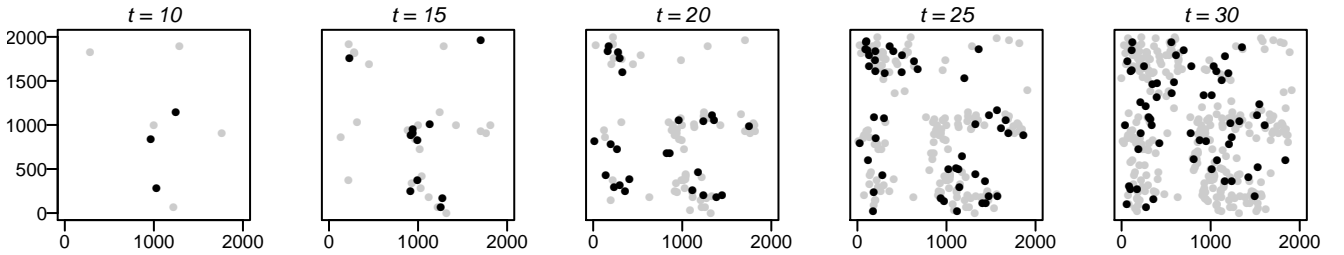


Figure 1: An illustration of (a subset of) the observed data y in the simulation (replicate 1) in the 2000×2000 square area in the form of a sequence ‘snapshots’ of the system state at particular times. Black and grey dots represent the individuals in class I and R respectively at times $t = 10, 15, 20, 25, 30$. It is assumed that the locations of all other individuals are known but these are not shown in the interests of clarity.

To assess our model testing framework we fit to the simulated data y a model with the correct structure (Case A), and three further models in which the spatial kernels (Case B & Case C) and the latent period distribution (Case D) have been mis-specified respectively. Specifically we consider:

- Case A: $K(d_{ij}, \kappa_1) = \exp(-\kappa_1 d_{ij})$ and the latent period is distributed as $\text{Gamma}(\mu, \sigma^2)$;
- Case B: $K(d_{ij}, \kappa_2) = 1/(1 + d_{ij}/\kappa_2)$ and the latent period is distributed as $\text{Gamma}(\mu, \sigma^2)$;
- Case C: $K(d_{ij}, \kappa_3) = d_{ij}^{-\kappa_3}$ and the latent period is distributed as $\text{Gamma}(\mu, \sigma^2)$;
- Case D: $K(d_{ij}, \kappa_1) = \exp(-\kappa_1 d_{ij})$ and the latent period is distributed as $\text{Exp}(\mu)$.

A *Weibull* infectious period is fitted in all cases. Uniform priors, which should be constrained to bounded regions to ensure a proper posterior distribution, are specified for all model parameters estimated in the following analyses.

In each case, we use standard MCMC and data augmentation to generate a sample from $\pi(\theta, \mathbf{z}|y)$ from which - see section 2 - we impute posterior samples of the Infection-Link Residuals (ILR, \tilde{r}_2) and of the Latent Time Residuals (LTR, \tilde{r}_3). In addition we impute posterior samples of \tilde{r}_1 . The *Anderson-Darling* test for consistency

Table 1: Values of $\pi(P(\tilde{r}_j) < 0.05|y)$, $j = 1, 2, 3$, estimated from 1,500 posterior samples of the corresponding components of $\tilde{\mathbf{r}}$ for 3 replicate epidemics simulated from the specified model and analysed using four different model assumptions (Case A, the correct model structure; Case B, a mis-specified (Cauchy-type) spatial kernel; Case C, a mis-specified (power-law) spatial kernel; Case D, a mis-specified latent period distribution).

	$\pi(P(\tilde{r}_1) < 0.05 y)$			$\pi(P(\tilde{r}_2) < 0.05 y)$			$\pi(P(\tilde{r}_3) < 0.05 y)$	
	Case B	Case C	Case D	Case A	Case B	Case C	Case A	Case D
Replicate 1	8%	5%	3%	6%	100%	80%	4%	99%
Replicate 2	5%	5%	4%	6%	100%	75%	4%	97%
Replicate 3	6%	6%	4%	5%	100%	76%	5%	100%

with the uniform distribution is applied to each sample of the residuals. Posterior distributions of the model parameters are described in detail in the *ESM*.

Table 1 shows the values of $\pi(P(\tilde{r}_j) < 0.05|y)$, $j = 1, 2, 3$, from three independently simulated replicates, y , of the epidemic. From $\pi(P(\tilde{r}_2) < 0.05|y)$ and $\pi(P(\tilde{r}_3) < 0.05|y)$ it appears that these posterior summaries systematically give evidence against the model when the spatial kernel and the latent period have been mis-specified respectively. On the other hand $\pi(P(\tilde{r}_1) < 0.05|y)$ suggests no evidence against the model specifications in Cases B, C and D. Note that, for ease of comparison, we only present the value of $\pi(P(\tilde{r}_j) < 0.05|y)$ for relevant cases. Values in cases not presented range from 3% to 6% and, therefore, suggest no evidence against the respective model specification.

3.1 Diagnosing Model Mis-specification

We now illustrate the insights our approach offers for understanding the causes of model inadequacy. Specifically, having observed considerable evidence against a model from the measure $\pi(P(\tilde{\mathbf{r}}) < 0.05|y)$, we show that the pattern of residuals $\pi(\tilde{r}_2|y)$ can suggest the manner in which the fitted model may be deficient. Consider two (symmetric) scenarios. In Scenario I, the epidemic is simulated from a kernel $K(d, \kappa_2) = d^{-2.8}$ and fitted with a kernel $K(d, \kappa_1) = \exp(-\kappa_1 d)$; in Scenario II, the epidemic is simulated from a kernel $K(d, \kappa_1) = \exp(-0.03d)$ and fitted with a kernel $K(d, \kappa_2) = d^{-k_2}$. Under the assumption that the fitted model is correct the imputed ILR should resemble samples from $U(0, 1)$. To highlight any systematic deviations from this null hypothesis Figure 2 presents the histogram formed from the subset of the ILR processes that produce small p-values (< 0.05) revealing a symmetry between the two scenarios. Scenario I and Scenario II respectively lead to a concentration or a scarcity of residuals at the extremes of the unit interval. This symmetry suggests that nature of the incompatibility of the spatial kernel may be diagnosed from systematically different deviations of the distribution of the ILR from $U(0, 1)$. This is discussed in more detail in *ESM* (Section 1.4).

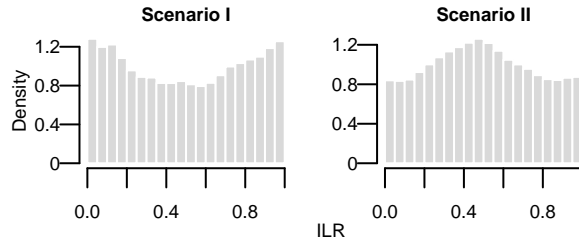


Figure 2: The distributions of a subset of imputed \tilde{r}_2 whose $P(\tilde{r}_2) < 0.05$ under two scenarios.

3.2 Comparison with Common Bayesian Model Checking Techniques

One common tool for model checking is the *Deviance Information Criterion* (*DIC*) [17]. The model with smallest DIC corresponds to the best model and, conventionally, models whose DIC exceeds that of the best model by more than 10 units are considered to display substantial evidence of poor fit. A key limitation of this approach is that it is known to be problematic when applied to processes that are only partially observed, as in the case of most real-world systems, where the DIC cannot be uniquely defined [18]. Following [18], we compute two versions of

DIC,

$$DIC_1 = -4\mathbb{E}_{\boldsymbol{\theta}, \mathbf{X}}[\log(f(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta})|\mathbf{y})] + 2\mathbb{E}_{\mathbf{X}}[\log(f(\mathbf{y}, \mathbf{X}|\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}])|\mathbf{y})] \quad (11)$$

and

$$DIC_2 = -4\mathbb{E}_{\boldsymbol{\theta}, \mathbf{X}}[\log(f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})|\mathbf{y})] + 2\mathbb{E}_{\mathbf{X}}[\log(f(\mathbf{y}|\mathbf{X}, \hat{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{X}))|\mathbf{y})] \quad (12)$$

where \mathbf{X} and \mathbf{y} represent the unobserved and observed data respectively and $\hat{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{X})$ is often estimated by posterior point estimates such as the posterior mean which is used here (note that DIC_1 and DIC_2 are referred to as DIC_4 and DIC_8 in [18]). The quantities $f(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta})$ and $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$ represent contributions to the likelihood from both the observation model and the process model in the first case, and the observation model alone in the second. Notice that calculation of each version of the DIC requires expectations of these quantities which can be estimated using MCMC techniques. The main difference between DIC_1 and DIC_2 is that the first takes the unobserved data into account. However, there is no absolute theoretical justification for a preference of one definition over another.

Table 2 shows that, although both versions of DIC can differentiate the *relative* goodness-of-fit between Case A and Case C as well as that between Case A and Case D, DIC_2 misleadingly suggests that the fit for Case B is better than that for Case A. Note that the DIC is not a direct measure of model adequacy and only measures the relative goodness-of-fit between two models. Moreover, as shown in Table 2, the ranking of models can also vary between different versions of DIC.

Table 2: DIC computed for Cases A, B, C and D

Replicate 1	Case A	Case B	Case C	Case D
DIC_1	10357.52	11561.90	10542.75	11525.69
DIC_2	5754.594	4982.372	5937.58	6897.95

A further approach to model checking is to compare the posterior predictive distribution of summary statistics with their observed values [20, 28]. In the *ESM* (Section 6), we consider posterior predictive checks based on common spatial autocorrelation measures including Moran’s I and Geary’s c indices [33], where application to simulated data shows that these measures are insensitive to the choice of model. Such summary statistics are based only on partially observed data and therefore reflect the behaviour of the competing models averaged over the missing data. In contrast, our method is based on the full posterior distribution of (imputed) unobserved data and other model parameters, which may explain its higher sensitivity to model mis-specification.

We further consider the performance of DIC and posterior predictive checks in an application to British floristic atlas data in Section 4.

4 Case Study: Spread of Giant Hogweed in Great Britain

Invasive alien species represent a major threat to ecosystems and cause significant environmental and financial loss worldwide [34]. *Heracleum mantegazzianum* (giant hogweed) causes significant problems in Great Britain and has rapidly spread since 1970 [27]. We apply our testing framework to British floristic atlas data which assess the presence of giant hogweed over a square lattice of 10×10 km resolution in 1970, 1987 and 2000. In total 2,838 such squares are considered to be habitable for the giant hogweed (see [27]). These are classified as susceptible or colonised at the given survey times according to the absence or presence of giant hogweed in the lattice. These data are well suited to testing our methodology. Detection of the species is relatively easy due to its height (> 2m), so that the number of false absences in the data set should be limited. Moreover, the data give ‘snapshots’ of the distribution at three distinct times (from 1970 to 2000) and over a large region making them particularly suitable for inferring the spatio-temporal transmission mechanism.

We first represent the lattice of square regions as a lattice of points where the position of a point is given by centre of the square which it represents. Figure 3 shows the snapshots of the spread of giant hogweed in Great Britain taken at three distinct times (1970, 1987 and 2000). In the light of the aggressive nature of giant hogweed we assume that, once colonised, sites can immediately start to colonise other sites and remain colonised. In the terminology of the epidemic model we consider, therefore, the E and I classes to be a single class and dispense with the recovery class R from our model. Effectively, we fit an S-I (Susceptible-Infectious) model to the presence/absence data and use our model assessment methods to compare the goodness-of-fit of several formulations, discussed in detail in the *ESM*. In summary, the models differ in the choice of spatial kernel and with regard to the inclusion of terms quantifying the suitability of each site, j , for the species. Suitability is represented by a measure $c_j \in [0, 1]$, where the c_j are

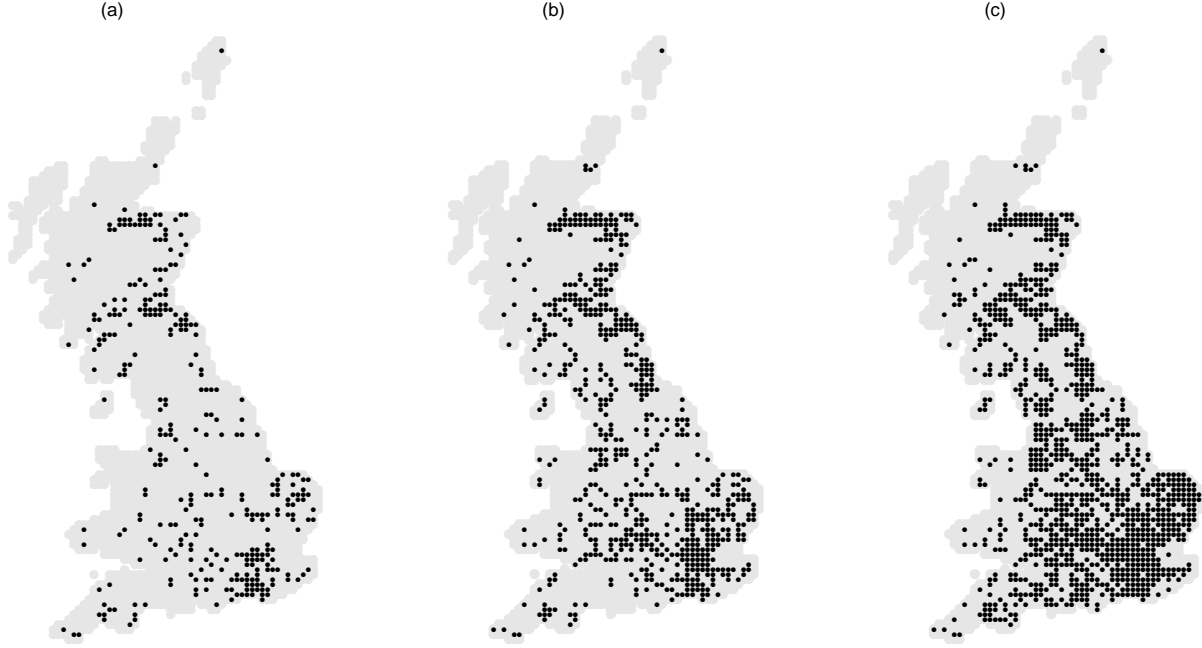


Figure 3: Snapshots of the spread of giant hogweed in Great Britain taken at three distinct times: (a) 1970, (b) 1987 and (c) 2000. Black dots represent the colonised sites.

taken from an earlier analysis [27] in which an extensive range of covariates including average temperature, altitude and other factors, were considered in their estimation. With suitability included, the instantaneous rate at which a susceptible site j becomes colonised Equation (1) is moderated by a factor c_j .

Full model specifications and posterior estimates of the model parameters are described in the *ESM*. Specifically, we consider three forms of spatial kernel with and without homogeneous suitabilities giving rise to six models:

- Model 1 (M1, Kernel A): $K(d_{ij}, \kappa_1) = \exp(-\kappa_1 d_{ij})$, heterogeneous suitabilities, c_j ;
- Model 2 (M2, Kernel B): $K(d_{ij}, \kappa_2) = 1/(1 + d_{ij}/\kappa_2)$, heterogeneous suitabilities, c_j ;
- Model 3 (M3, Kernel C): $K(d_{ij}, \kappa_3) = d_{ij}^{-\kappa_3}$, heterogeneous suitabilities, c_j ;
- Model 4 (M4, Kernel A): $K(d_{ij}, \kappa_1) = \exp(-\kappa_1 d_{ij})$, homogeneous suitabilities $c_j = 1$;
- Model 5 (M5, Kernel B): $K(d_{ij}, \kappa_2) = 1/(1 + d_{ij}/\kappa_2)$, homogeneous suitabilities $c_j = 1$;
- Model 6 (M6, Kernel C): $K(d_{ij}, \kappa_3) = d_{ij}^{-\kappa_3}$, homogeneous suitabilities $c_j = 1$.

These models are fitted to the data using Bayesian methods as described in the *ESM*. For the simple SI formulation the residual process reduces to $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2)$ and specifies exposure times and infection links. We apply three tests to imputed values of these residuals for each of the models. As with the simulated example, we investigate $\pi(P(\tilde{r}_1) < 0.05|y)$ and $\pi(P(\tilde{r}_2) < 0.05|y)$ arising from an Anderson-Darling test applied to the respective subset of residuals. We also consider a combined test, with p-value $P(\tilde{\mathbf{r}})$, based on a test statistic

$$T(\tilde{\mathbf{r}}) = -2(\log(P(\tilde{r}_1)) + \log(P(\tilde{r}_2)))$$

whose distribution under the modelling assumptions is χ_4^2 , and report $\pi(P(\tilde{\mathbf{r}}) < 0.05|y)$ for each model. Conclusions arising from the various tests are now presented.

4.1 Model Assessment and Implications for Control Strategies

From Table 3 we first notice that, regardless of whether dependence on suitability is included, $\pi(P(\tilde{r}_2) < 0.05|y)$ and $\pi(P(\tilde{\mathbf{r}}) < 0.05|y)$ is larger for the models with Cauchy-form kernel (Kernel B, M2 & M5) than for the respective models with exponentially bounded kernel (Kernel A, M1 & M4) or with power-law kernel (Kernel C, M3 & M6),

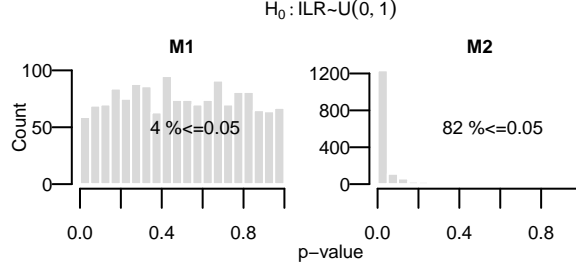


Figure 4: Posterior distributions of the p-values from testing the sets of posterior samples of Infection Link Residual (ILR) imputed from MCMC chains (1,500 samples in each case) when fitting SI models, representing heterogeneous suitability, to the giant hogweed data with kernel A (model M1) and kernel B (model M2) respectively

Table 3: Values of $\pi(P(\tilde{r}_j) < 0.05|y)$, $j = 1, 2$, and $\pi(P(\tilde{r}) < 0.05|y)$ estimated from 1,500 posterior samples of ILR and ETR computed from the giant hogweed data under different model assumptions regarding the spatial kernel and dependence on suitability of sites

Spatial Kernels	$\pi(P(\tilde{r}_1) < 0.05 y)$			$\pi(P(\tilde{r}_2) < 0.05 y)$			$\pi(P(\tilde{r}) < 0.05 y)$		
	A	B	C	A	B	C	A	B	C
Heterogeneous Suitability	13%	7%	11%	4%	82%	4%	11%	74%	10%
Homogeneous Suitability	5%	6%	3%	35%	100%	54%	24%	100%	28%

suggesting that the Cauchy kernel typically provides a poorer fit. When dependence on suitability is not included (M4, M5 & M6), the fact $\pi(P(\tilde{r}_2) < 0.05|y) = 0.35$ for M4 (Kernel A) and $\pi(P(\tilde{r}_2) < 0.05|y) = 0.54$ for M6 (Kernel C) suggests there are substantial probabilities that the $U(0,1)$ hypothesis for the imputed residuals would be rejected by the classical observer and calls these models into question. By contrast, the results for M1, M2 and M3 present no evidence against the model with exponentially bounded kernel (A) and power-law kernel (C), while there remains a substantial posterior probability of rejection (0.82) for the model with Cauchy-form kernel (B). Figure 4 presents samples from $\pi(P(\tilde{r}_2)|y)$ for M1 and M2, highlighting the evidence against kernel B. It is clear from other results in Table 3 that the evidence against a given model arises from the ILR residuals \tilde{r}_2 ; the test based on \tilde{r}_1 alone presents little evidence against any of the models M1-M6, while the evidence arising from the combined test is typically weaker than that from the tests of \tilde{r}_2 alone.

In summary we find evidence that the dispersion mechanism for hogweed cannot be represented adequately by the Cauchy dispersal kernel, while no evidence against the exponentially-bounded kernel and power-law kernel is found as long as habitat heterogeneity is accommodated. Figure S8 (ESM) shows that the long-tail behaviour of M2 tends to induce a scarcity of residuals at the left end and a concentration of residuals at the right end of the unit interval. Although the ILR residuals \tilde{r}_2 were constructed with assessment of spatial kernels in mind, comparison of $\pi(P(\tilde{r}_2) < 0.05|y)$ between models with and without the dependence on suitability (i.e., comparison between M1 and M4 and that between M3 and M6) highlights the potential for the method to detect mis-specification of other aspects of the model. This is not surprising given the key role of \tilde{r}_2 and the suitabilities in the construction of the colonisation links.

Giant hogweed spread their seed mostly through wind, water and human activities [35]. Localized dispersal mechanisms typically involve the dispersal by wind or animal activities. Human activities, such as soil transport and transport of seeds adhering to the car tyres, are mainly responsible for long-distance dispersal. Understanding of the importance of short-distance and long-distance dispersal provides valuable insight for devising appropriate control strategies [36, 37]. Our results and Figure 5 clearly suggest that the spread of hogweed is mainly via a nearest-neighbour mechanism. Given this highly localized dispersal mechanism, and the constraints imposed by the lattice structure of the hogweed data (Section 4) which forces a minimum distance of 10km between two sites (in contrast to a more general continuous space in the simulated example), the difference between an exponentially-bounded kernel and a power-law kernel becomes insignificant (also see Figure 5). Hence the two models display similar goodness-of-fit. This suggests that control measures - for instance, education programs of increasing public awareness and participation in prevention and reporting [38], and field survey and subsequent eradication measures [35, 39] - may be most effectively deployed by focusing implementation on the neighbourhood of known colonisations.

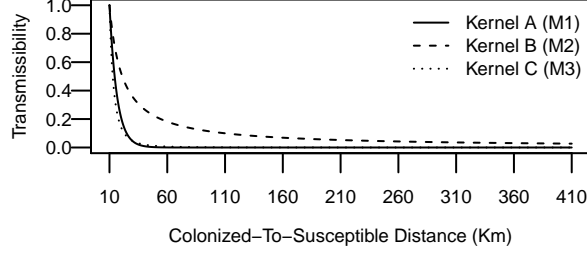


Figure 5: Estimated spatial kernels from fitting M1 and M2 and M3 to the giant hogweed data with kernel parameters set to posterior means. Transmissibility is expressed relative to the amplitude of the respective kernel at 10km (the minimum distance between two sites in the data set)

4.2 Comparison with DIC and Posterior Predictive Checks

Similar to Section 3, we compute two versions of DIC, DIC_1 and DIC_2 (see Equation (11) and Equation (12)), for M1 and M2 and the corresponding models M4 and M5 which do not take the suitability into account. From Table 4, we note that, although both versions of DIC can differentiate the *relative* goodness-of-fit between M1 and M2, they unreasonably indicate that M4 (exponentially-bounded kernel without considering suitability) is the best or the second best model.

In Section 3 and in the *ESM* (Section 6), we have shown that summary statistics quantifying spatial autocorrelation may be less sensitive to model mis-specification than our latent residual approach. We therefore focus on more intuitive summary statistics based on the number of colonisations. We also adopt a conservative approach, running forward simulations of the fitted model using point estimates of parameters, here the posterior mean. Moreover, these simulations are conditioned on the colonised sites observed in 1970. We focus on comparing models M1 and M2 (which include dependence on suitability and for which our residual analysis shows a significant difference in goodness-of-fit) and examine the following predictive outcomes: the predictive distribution of the number of colonized sites, at the end of the observation period (2000), within annular regions centered on a given location (see Figure S7 in ESM for a representation of the regions), and the numbers of reported colonized sites at the second and third observation times (1987 and 2000). Figure 6 and Table 5 compare predicted distributions and the actual observations.

Table 4: DIC computed for M1, M2, M4 and M5.

Model	M1	M2	M4	M5
DIC_1	7404.8	7442.9	7422.0	7890.1
DIC_2	968.3	1027.4	522.2	1194.8

Table 5: Predicted and reported new colonized sites at second and third snapshots (1987 and 2000). The reported numbers are followed by the two-sided 95% credible intervals enclosed in brackets (1,000 simulations for each model)

Model	M1	M2
1987	334 (311, 375)	334 (310, 381)
2000	412 (368, 434)	412 (388, 460)

Figure 6 and Table 5 suggest that, similar to Section 3.2, model checks based on these apparently reasonable summary statistics may be insensitive to the choice of model. Figure 5 shows that M1 and M2 represent very different transmission mechanisms, with kernel B exhibiting a strong propensity for long-range transmission. Figure 7 shows that the estimates of transmission rates can be different when different kernels are fitted. Nevertheless, the predictive distributions of the summary statistic appear consistent with observed values for both M1 and M2.

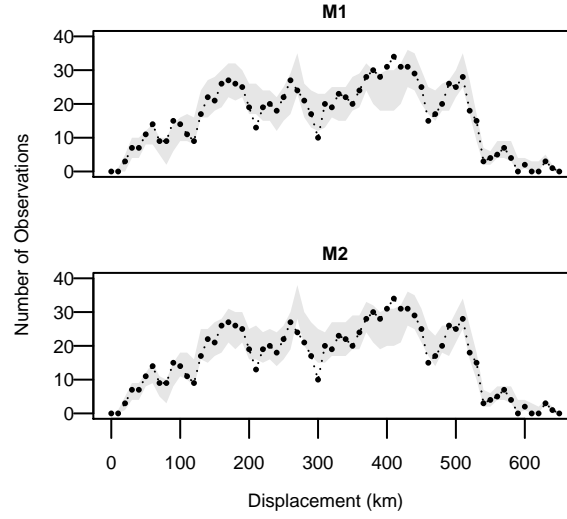


Figure 6: Distribution of the number of colonized sites within each ring region at the final observation time as predicted by models (the shaded area represents the 95% two-sided interval of the predicted number of colonized sites from 1,000 simulations) and the observed data (shown in the dotted black line). The displacements are measured from the center of the ring regions

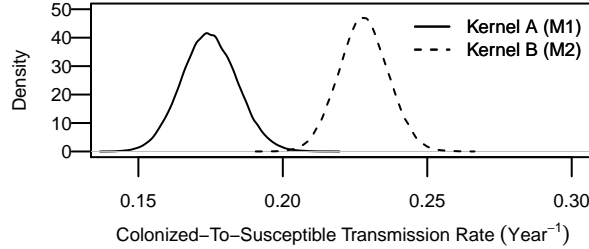


Figure 7: Posterior distributions of transmission rates from a colonized site to a susceptible site from fitting M1 and M2 to the giant hogweed data

5 Discussion

Earlier work [40] has championed the view that Bayesian and classical reasoning are natural approaches to follow for parameter estimation and model criticism respectively and therefore should be used in combination. We have presented a statistical framework that combines classical and Bayesian reasoning in testing for mis-specifications of a spatio-temporal model by investigating the consistency of imputed latent residuals with a known sampling distribution using a classical hypothesis test. In particular, we have shown how such residuals can be tailored so as to be sensitive to mis-specifications of the spatial kernel or the sojourn-time distributions within a spatio-temporal model. Analysis of simulated epidemics and data on the spread of an invasive species in the UK demonstrates how our model testing framework can be utilized to diagnose the model fit and how the results can be interpreted in practice. In Section 3.1, we also discuss how one might perform the complementary diagnosis of the type of incompatibility of a spatial kernel by examining deviations (from the null hypothesis that the fitted model is correct) in the distribution of residuals $\pi(\tilde{\tau}_2|y)$ after observing considerable evidence against the model.

Predicted dynamics of epidemiological and ecological systems can be extremely sensitive to the choice of model, with consequent implications for the design of control strategies [1, 5, 6, 7]. For example, choices of temporal distributions influence both epidemic final size and the persistence of disease, and have important implications for devising effective control strategies which target symptomatic subjects and the timing of infectiousness [6, 12, 13]. There are examples of such effects related to the parametric form of incubation and infectious period distributions in models of measles [9, 10, 11], smallpox [12] and AIDS [13]. It is well-known that the spatial transmission mechanisms are difficult to assess in practice yet have major implications for optimal control strategies. Studies

of animals and plant diseases such as foot and mouth and citrus canker have cited the importance for selecting between a long-tailed spatial kernel versus a localized spatial kernel when devising the most appropriate strategy of culling [41, 8, 5]. Therefore we believe that the methodology presented here, based on infection-link residuals (ILR), is a novel and potentially powerful tool for diagnosing mis-specification of a spatial kernel which can provide valuable insights to modellers in practice. Moreover, we remark that the principles introduced here should be readily extendable allowing the construction of analogous residuals for a wide range of processes included in models in ecology and epidemiology.

We also believe the approach offers several advantages over alternatives. Bayesian model assessment approaches, such as Bayes factors, are known to be sensitive to selection of prior distributions and are challenging computationally [42, 43]. Moreover, they allow only relative comparison of competing models, a disadvantage shared by information criteria measures such as the Deviance Information Criterion (DIC) [17]. The latter is also problematic when dealing with partially observed processes [18], the norm in epidemiological studies, where the DIC is not uniquely defined. By contrast, the tests based on latent residuals offer an assessment of model discrepancy in absolute terms. Posterior predictive checks that utilize only partially observed data may be insensitive to the model choice (as shown in section 3 and section 4.2) even if summary statistics are appropriately chosen. A key feature of the proposed tests is that they can be easily embedded within any Bayesian analysis of a spatio-temporal system that makes use of data augmentation. Also, in contrast to other approaches such as posterior predictive checks, our method utilizes the full posterior distribution of unobserved data and model parameters, and may offer a higher sensitivity to model mis-specifications. As it is common practice to conduct Bayesian analyses of partially observed epidemics using data augmentation supported by computational techniques such as Markov chain Monte Carlo methods [5, 27, 28, 20], the framework represents a potentially valuable addendum to the model-testing toolkit used in epidemiological and ecological studies.

Acknowledgements

We acknowledge financial support from Heriot-Watt University and the Scottish Government’s Rural and Environment Science and Analytical Services Division (RESAS). The distribution data were obtained from the National Biodiversity Network Gateway, and are compiled from numerous sources including the Countryside Council for Wales, Bristol Regional Environmental Records Centre, The Scottish Wildlife Trust, and Scottish Borders Biological Records Centre (for details see www.nbn.org.uk/). We also thank Dr Stephen Catterall for helping with the hogweed data and providing suitability estimates used in section 4. Finally, we are thankful to the Editor and three referees for their constructive reviews of the manuscript, which have helped us to introduce some further insights into the paper.

References

- [1] Ster IC, Singh BK, Ferguson NM (2009) Epidemiological inference for partially observed epidemics: the example of the 2001 foot and mouth epidemic in Great Britain. *Epidemics* 1:21–34. (doi: 10.1016/j.epidem.2008.09.001)
- [2] Boender GJ, Meester R, Gies E, De Jong M (2007) The local threshold for geographical spread of infectious diseases between farms. *Preventive Veterinary Medicine* 82:90–101. (doi: 10.1016/j.prevetmed.2007.05.016)
- [3] Cook AR, Marion G, Butler A, Gibson GJ (2007) Bayesian inference for the spatio-temporal invasion of alien species. *Bulletin of Mathematical Biology* 69:2005–2025. (doi: 10.1007/s11538-007-9202-4)
- [4] Walters RJ, Hassall M, Telfer MG, Hewitt GM, Palutikof JP (2006) Modelling dispersal of a temperate insect in a changing climate. *Proceedings of the Royal Society B: Biological Sciences* 273:2017–2023. (doi:10.1098/rspb.2006.3542)
- [5] Ferguson NM, Donnelly CA, Anderson RM (2001) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* 292:1155–1160. (doi: 10.1126/science.1061020)
- [6] Fraser C, Riley S, Anderson RM, Ferguson NM (2004) Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the United States of America* 101(16):6146–6151. (doi: 10.1073/pnas.0307506101)
- [7] Filipe JAN, Maule MM (2004) Effects of dispersal mechanisms on spatio-temporal development of epidemics. *Journal of Theoretical Biology* 226:125–141.

- [8] Keeling MJ, Woolhouse MEJ, May RM, Davies G, Grenfell BT (2003) Modelling vaccination strategies against foot-and-mouth disease. *Nature* 421:136–142. (doi: 10.1038/nature01343)
- [9] Ferguson NM, May RM, Anderson RM (1997) Measles: Persistence and synchronicity in disease dynamics. *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions* 30:137–157. Princeton University Press Princeton
- [10] Bolker B, Grenfell B (1995) Space, persistence and dynamics of measles epidemics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 348:309–320. (doi: 10.1098/rstb.1995.0070)
- [11] Keeling MJ, Grenfell BT (1997) Disease extinction and community size: modeling the persistence of measles. *Science* 275(5296):65–67. (doi: 10.1126/science.275.5296.65)
- [12] Streftaris G, Gibson GJ (2004) Bayesian inference for stochastic epidemics in closed populations. *Statistical Modelling* 4:63–75. (doi: 10.1191/1471082X04st065oa)
- [13] Muñoz A, Sabin C, Phillips A (1997) The incubation period of AIDS. *Aids* 11:S69–S76.
- [14] Jeffreys H (1935) *Some tests of significance, treated by the theory of probability*. Proceedings of the Cambridge Philosophical Society (Cambridge University Press), 31:203–222.
- [15] Jeffreys H (1961) *Theory of Probability* (Oxford University Press, USA).
- [16] Draper D (1995) Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 45–97.
- [17] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:583–639. (doi: 10.1111/1467-9868.00353)
- [18] Celeux G, Forbes F, Robert CP, Titterton DM (2006) Deviance information criteria for missing data models. *Bayesian Analysis* 1(4):651–673 (doi: 10.1214/06-BA122)
- [19] Meng XL (1994) Posterior predictive p-values. *The Annals of Statistics* 22(3): 1142–1160. (doi: 10.1214/aos/1176325622)
- [20] Gibson GJ, Otten W, Filipe JA, Cook A, Marion G, Gilligan CA (2006) Bayesian estimation for percolation models of disease spread in plant populations. *Statistics and Computing* 16:391–402. (doi: 10.1007/s11222-006-0019-z)
- [21] Papaspiliopoulos O, Roberts GO, Sköld M (2007) A general framework for the parametrization of hierarchical models. *Statistical Science* 22:59–73. (doi: 10.1214/088342307000000014)
- [22] Streftaris G, Gibson GJ (2004) Bayesian analysis of experimental epidemics of foot-and-mouth disease. *Proceedings of the Royal Society of London-B* 271:1111–1118. (doi: 10.1098/rspb.2004.2715)
- [23] Valleron AJ, Boelle PY, Will R, Cesbron JY (2001) Estimation of epidemic size and incubation time based on age characteristics of vcjd in the united kingdom. *Science* 294:1726–1728. (doi: 10.1126/science.1066838)
- [24] Anderson RM (1988) The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *Journal of the Royal Statistical Society: Series B* 151:66–93. (doi: 10.2307/2982185)
- [25] Dawid AP, Stone M (1982) The functional-model basis of fiducial inference. *The Annals of Statistics* 10(4): 1054–1067. (doi: 10.1214/aos/1176345970)
- [26] Cox DR, Snell EJ (1968) A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)* 30(2): 248–275.
- [27] Catterall S, Cook AR, Marion G, Butler A, Hulme PE (2012) Accounting for uncertainty in colonisation times: a novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. *Ecography* 35(10):901–911. (doi: 10.1111/j.1600-0587.2011.07190.x)
- [28] Cook AR, Otten W, Marion G, Gibson GJ, Gilligan CA (2007) Estimation of multiple transmission rates for epidemics in heterogeneous populations. *Proceedings of the National Academy of Sciences* 104:20392–20397.

- [29] Lewis PAW (1961) Distribution of the Anderson-Darling statistic. *The Annals of Mathematical Statistics* 32(4): 1118–1124 (doi: 10.1214/aoms/1177704850)
- [30] Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association* 88:669–679. (doi: 10.1080/01621459.1993.10476321)
- [31] Sellke T (1983) On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability* 20(2): 390–394. (doi: 10.2307/3213811)
- [32] Streftaris G, Gibson GJ (2012) Non-exponential tolerance to infection in epidemic systems - modeling, inference, and assessment. *Biostatistics* 13:580–593. (doi: 10.1093/biostatistics/kxs011)
- [33] Getis A (1991) Spatial interaction and spatial autocorrelation: a cross-product approach. *Environment and Planning A* 23(9):1269–1277.
- [34] Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the united states. *Ecological Economics* 52:273–288. (doi: 10.1016/j.ecolecon.2004.10.002)
- [35] Pyšek P (2007) *Ecology and management of giant hogweed:(Heracleum mantegazzianum)*. CABI
- [36] Dawe NK , White ER (1979) Giant cow parsnip (*Heracleum mantegazzianum*) on Vancouver Island, British Columbia. *Canadian Field-Naturalist* 93(1):82–83
- [37] Pergl J, Müllerová J, Perglová I, Herben I, Pyšek P (2011) The role of long-distance seed dispersal in the local population dynamics of an invasive plant species. *Diversity and Distributions* 17(4):725–738 *Wiley Online Library* (doi: 10.1111/j.1472-4642.2011.00771.x)
- [38] Bhowmik, PC: Characteristics, significance, and human dimension of global invasive weeds. In: *Invasive plants: ecological and agricultural aspects*, pp. 251–268. Indejit (ed.) Springer, Alemania (2005)
- [39] Sampson C, Waal LD., Child LE, Wade PM, Brock JH (1994) Cost and impact of current control methods used against *Heracleum mantegazzianum* (giant hogweed) and the case for instigating a biological control programme. *Ecology and Management of Invasive Riverside Plants* pp. 55-65 John Wiley & Sons Ltd
- [40] Box GEP (1980) Sampling and bayes’ inference in scientific modelling and robustness. *Journal of the Royal Statistical Society: Series A (General)* 143(4): 383–430. (doi: 10.2307/2982063)
- [41] Gottwald TR, Graham JH, Schubert TS (2002) Citrus canker: the pathogen and its impact. Online. *Plant Health Progress* (doi:10.1094/PHP-2002-0812-01-RV)
- [42] Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* 90(430):773–795.
- [43] Han C, Carlin BP (2001) Markov chain Monte Carlo methods for computing Bayes factors. *Journal of the American Statistical Association* 96(455):1122–1132.